(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 024 477 A1

EP 18741

(12) EUROPEAN PATENT APPLICATION
published in accordance with Art. 158(3) EPC

(43) Date of publication:
02.08.2000 Bulletin 2000/31

(21) Application number: 99940456.9

(22) Date of filing: 20.08.1999

(51) Int. Cl.7: G10L 11/00

(86) International application number:
PCT/JP99/04468

(87) International publication number:
WO 00/11646 (02.03.2000 Gazette 2000/09)

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: 21.08.1998 JP 23614798
21.09.1998 JP 26688398

(71) Applicant:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka 571-8501 (JP)

(72) Inventor: EHARA, Hiroyuki
Yokohama-shi Kanagawa 233-0016 (JP)

(74) Representative:
Grünecker, Kinkeldey,
Stockmair & Schwanhäusser
Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(54) MULTIMODE SPEECH ENCODER AND DECODER

(57) Excitation information is coded in multimode using static and dynamic characteristics of quantized vocal tract parameters, and also at a decoder side, the postprocessing is performed in the multimode, thereby improving the qualities of unvoiced speech region and stationary noise region.

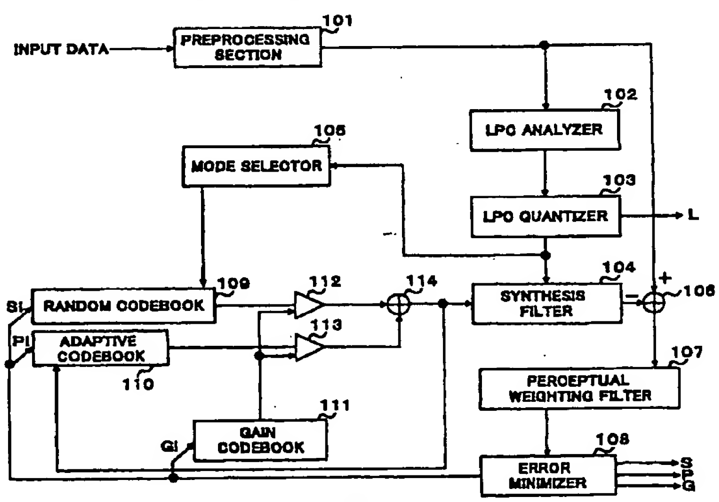

FIG.1

EP 1 024 477 A1

FIG.11 is a flowchart for the former part of the multimode postprocessing in the fifth embodiment of the present invention; and

FIG.12 is a flowchart for the latter part of the multimode postprocessing in the fifth embodiment of the present invention.

Best Mode for Carrying Out the Invention

[0009] Speech coding apparatuses and others in embodiments of the present invention are explained below using FIG.1 to FIG.9.

(First embodiment)

[0010] FIG.1 is a block diagram illustrating a configuration of a speech coding apparatus according to the first embodiment of the present invention.

[0011] Input data, comprised of, for example, digital speech signals, is input to preprocessing section 101. Preprocessing section 101 performs processing such as cutting of a direct current component and bandwidth limitation of the input data using a high-pass filter and band-pass filter to output to LPC analyzer 102 and adder 106. In addition, although it is possible to perform successive coding processing without performing any processing in preprocessing section 101, the coding performance is improved by performing the above-mentioned processing.

[0012] LPC analyzer 102 performs linear prediction analysis, and calculates linear predictive coefficients (LPC) to output to LPC quantizer 103.

[0013] LPC quantizer 103 quantizes the input LPC, outputs the quantized LPC to synthesis filter 104 and mode selector 105, and further outputs a code L that represents the quantized LPC to decoder. In addition, the quantization of LPC is performed usually after LPC is converted to LSP (Line Spectrum Pair) which has better interpolation characteristics.

[0014] As synthesis filter 104, a LPC synthesis filter is constructed using the quantized LPC input from LPC quantizer 103. With the constructed synthesis filter, filtering processing is performed on an excitation vector signal input from adder 114, and the resultant signal is output to adder 106.

[0015] Mode selector 105 determines a mode of random codebook using the quantized LPC input from LPC quantizer 103.

[0016] At this time, mode selector 105 stores previously input information on quantized LPC, and performs the selection of mode using both characteristics of an evolution of quantized LPC between frames and of the quantized LPC in a current frame. There are at least two types of the modes, of which examples are a mode corresponding to a voiced speech segment, and a mode corresponding to an unvoiced speech segment and stationary noise segment. Further, as information for use in selecting a mode, it is not necessary to use the quantized LPC themselves, and it is more effective to use converted parameters such as the quantized LSP, reflective coefficients and linear prediction residual power.

[0017] Adder 106 calculates an error between the preprocessed input data input from preprocessing section 101 and the synthesized signal to output to perceptual weighting filter 107.

[0018] Perceptual weighting filter 107 performs perceptual weighting on the error calculated in adder 106 to output to error minimizer 108.

[0019] Error minimizer 108 adjusts a random codebook index Si, adaptive codebook index (pitch period) Pi, and gain codebook index Gi respectively output to random codebook 109, adaptive codebook 110, and gain codebook 111, determines a random code vector, adaptive code vector, and random codebook gain and adaptive codebook gain respectively to be generated in random codebook 109, adaptive codebook 110, and gain codebook 111 so as to minimize the perceptual weighted error input from perceptual weighting filter 107, and outputs a code S representing the random code vector, a code P representing the adaptive code vector, and a code G representing gain information to decoder.

[0020] Random codebook 109 stores the predetermined number of random code vectors with different shapes, and outputs the random code vector designated by the index Si of random code vector input from error minimizer 108. Random codebook 109 has at least two types of modes. For example, random codebook 109 is configured to generate a pulse-like random code vector in the mode corresponding to a voiced speech segment, and further generate a noise-like random code vector in the mode corresponding to an unvoiced speech segment and stationary noise segment. The random code vector output from random codebook 109 is generated with a single mode selected in mode selector 105 from among at least two types of the modes described above, and multiplied by the random codebook gain Gs in multiplier 112 to be output to adder 114.

[0021] Adaptive codebook 110 performs buffering while updating the previously generated excitation vector signal sequentially, and generates the adaptive code vector using the adaptive codebook index (pitch period (pitch lag)) input from error minimizer 108. The adaptive code vector generated in adaptive codebook 110 is multiplied by the adaptive codebook gain Ga in multiplier 113, and then output to adder 114.

[0022] Gain codebook 111 stores the predetermined number of sets of the adaptive codebook gain Ga and random codebook gain Gs (gain vector), and outputs the adaptive codebook gain component Ga and random codebook gain component Gs of the gain vector designated by the gain codebook index Gi input from error minimizer 108 respectively to multipliers 113 and 112. In addition, if the gain codebook is constructed with

has at least two types of the modes. For example, the search is performed by using the random codebook storing pulse-like random code vectors in the mode corresponding to the voiced speech segment, and using the random codebook storing noise-like random code vectors in the mode corresponding to the unvoiced speech segment and stationary noise segment. The random codebook of which mode is used in the search is selected in ST307.

[0035] Next, in ST310, gain codebook search is performed. The gain codebook search is to select from the gain codebook a pair of the adaptive codebook gain and random codebook gain respectively to be multiplied the adaptive code vector determined in ST308 and the random code vector determined in ST309. The excitation vector signal is generated by adding the adaptive code vector multiplied by the adaptive codebook gain and the random code vector multiplied by the random codebook gain. The pair of the adaptive codebook gain and random codebook gain is selected from the gain codebook so as to minimize an error between a signal obtained by filtering the generated excitation vector signal with the perceptual weighted synthesis filter constructed in ST306, and the signal obtained by filtering the preprocessed input data with the perceptual weighting filter constructed in ST305.

[0036] Next, in ST311, the excitation vector signal is generated. The excitation vector signal is generated by adding a vector obtained by multiplying the adaptive code vector selected in ST308 by the adaptive codebook gain selected in ST310 and a vector obtained by multiplying the random code vector selected in ST309 by the random Codebook gain selected in ST310.

[0037] Next, in ST312, the update of the memory used in a loop of the subframe processing is performed. Examples specifically performed are the update of the adaptive codebook, and the update of states of the perceptual weighting filter and perceptual weighted synthesis filter.

[0038] In ST305 to ST312, the processing is performed on a subframe-by-subframe basis.

[0039] Next, in ST313, the update of memory used in a loop of the frame processing. Examples specifically performed are the update of states of the filter used in the preprocessing section, the update of quantized LPC buffer (in the case where the inter-frame predictive quantization of LPC is performed), and the update of input data buffer.

[0040] Next, in ST314, coded data is output. The coded data is output to a transmission path while being subjected to bit stream processing and multiplexing processing corresponding to the form of the transmission.

[0041] In ST302 to 304 and ST313 to 314, the processing is performed on a frame-by-frame basis. Further the processing on a frame-by-frame basis and subframe-by-subframe is iterated until the input data is consumed.

(Second embodiment)

[0042] FIG.2 is a block diagram illustrating a configuration of a speech decoding apparatus according to the second embodiment of the present invention.

[0043] The code L representing quantized LPC, code S representing a random code vector, code P representing an adaptive code vector, and code G representing gain information, each transmitted from a coder, are respectively input to LPC decoder 201, random codebook 203, adaptive codebook 204 and gain codebook 205.

[0044] LPC decoder 201 decodes the quantized LPC from the code L to output to mode selector 202 and synthesis filter 209.

[0045] Mode selector 202 determines a mode for random codebook 203 and postprocessing section 211 using the quantized LPC input from LPC decoder 201, and outputs mode information M to random codebook 203 and postprocessing section 211. In addition, mode selector 202 also stores previously input information on quantized LPC, and performs the selection of mode using both characteristics of an evolution of quantized LPC between frames and of the quantized LPC in a current frame. There are at least two types of the modes, of which examples are a mode corresponding to a voiced speech segment, a mode corresponding to an unvoiced speech segment, and a mode corresponding to a stationary noise segment. Further, as information for use in selecting a mode, it is not necessary to use the quantized LPC themselves, and it is more effective to use converted parameters such as the quantized LSP, reflective coefficients and linear prediction residual power.

[0046] Random codebook 203 stores the predetermined number of random code vectors with different shapes, and outputs a random code vector designated by the random codebook index obtained by decoding the input code S. This random codebook 203 has at least two types of the modes. For example, random codebook 203 is configured to generate a pulse-like random code vector in the mode corresponding to a voiced speech segment, and further generate a noise-like random code vector in the modes corresponding to an unvoiced speech segment and steady noise segment. The random code vector output from random codebook 203 is generated with a single mode selected in mode selector 202 from among at least two types of the modes described above, and multiplied by the random codebook gain Gs in multiplier 206 to be output to adder 208.

[0047] Adaptive codebook 204 performs buffering while updating the previously generated excitation vector signal sequentially, and generates an adaptive code vector using the adaptive codebook index (pitch period (pitch lag)) obtained by decoding the input code P. The adaptive code vector generated in adaptive codebook 204 is multiplied by the adaptive codebook gain Ga in

[0063] Next, in ST409, the excitation vector signal is generated. The excitation vector signal is generated by adding a vector obtained by multiplying the adaptive code vector selected in ST406 by the adaptive codebook gain selected in ST408 and a vector obtained by multiplying the random code vector selected in ST407 by the random codebook gain selected in ST408.

[0064] Next, in ST410, a decoded signal is synthesized. The excitation vector signal generated in ST409 is filtered with the synthesis filter constructed in ST404, and thereby the decoded signal is synthesized.

[0065] Next, in ST411, the postfiltering processing is performed on the decoded signal. The postfiltering processing is comprised of the processing to improve subjective qualities of decoded signals, in particular, decoded speech signals, such as pitch emphasis processing, formant emphasis processing, spectral tilt compensation processing and gain adjustment processing.

[0066] Next, in ST412, the final postprocessing is performed on the decoded signal subjected to postfiltering processing. The postprocessing is comprised of the processing to improve subjective qualities of stationary noise segment in the decoded signal such as inter-(sub)frame smoothing processing of spectral amplitude and randomizing processing of spectral phase, and the processing corresponding to mode selected in ST405 is performed. For example, the smoothing processing and randomizing processing is rarely performed in the modes corresponding to the voiced speech segment and unvoiced speech segment, and such processing is performed in the mode corresponding to the stationary noise segment. The signal generated in this step becomes output data.

[0067] Next, in ST413, the update of the memory used in a loop of the subframe processing is performed. Specifically performed are the update of the adaptive codebook, and the update of states of filters used in the postfiltering processing.

[0068] In ST404 to ST413, the processing is performed on a subframe-by-subframe basis.

[0069] Next, in ST414, the update of memory used in a loop of the frame processing is performed. Specifically performed are the update of quantized (decoded) LPC buffer (in the case where the inter-frame predictive quantization of LPC is performed), and update of output data buffer.

[0070] In ST402 to 403 and ST414, the processing is performed on a frame-by-frame basis. Further, the processing on a frame-by-frame basis is iterated until the coded data is consumed.

(Third embodiment)

[0071] FIG.5 is a block diagram illustrating a speech signal transmission apparatus and reception apparatus respectively provided with the speech coding apparatus of the first embodiment 1 and speech decoding apparatus of the second embodiment 2. FIG.5A illustrates the transmission apparatus, and FIG.5B illustrates the reception apparatus.

[0072] In the speech signal transmission apparatus in FIG.5A, speech input apparatus 501 converts a speech into an electric analog signal to output to A/D converter 501. A/D converter 502 converts the analog speech signal into a digital speech signal to output to speech coder 503. Speech coder 503 performs speech coding processing on the input signal, and outputs coded information to RF modulator 504. R/F modulator 54 performs modulation, amplification and code spreading on the coded speech signal information to transmit as a radio signal, and outputs the resultant signal to transmission antenna 505. Finally, the radio signal (RF signal) 506 is transmitted from transmission antenna 505.

[0073] On the other hand, the reception apparatus in FIG.5b receives the radio signal (RF signal) 506 with reception antenna 507, and outputs the received signal to RF demodulator 508. RF demodulator 508 performs the processing such as code despreading and demodulation to convert the radio signal into coded information, and outputs the coded information to speech decoder 509. Speech decoder 509 performs decoding processing on the coded information and outputs a digital decoded speech signal to D/A converter 510. D/A converter 510 converts the digital decoded speech signal output from speech decoder 509 into an analog decoded speech signal to output to speech output apparatus 511. Finally, speech output apparatus 511 converts the electric analog decoded speech signal into a decoded speech to output.

[0074] It is possible to use the above-mentioned transmission apparatus and reception apparatus as a mobile station apparatus and base station apparatus in mobile communication apparatuses such as portable telephones. In addition, the medium that transmits the information is not limited to the radio signal described in this embodiment, and it may be possible to use optosignals, and further possible to use cable transmission paths.

[0075] Further, it may be possible to achieve the speech coding apparatus described in the first embodiment, the speech decoding apparatus described in the second embodiment, and the transmission apparatus and reception apparatus described in the third embodiment by recording the corresponding program in a recording medium such as a magnetic disk, optomagnetic disk, and ROM cartridge to use as software. The use of thus obtained recording medium enables a personal computer using such a recording medium to achieve the speech coding/decoding apparatus and transmission/reception apparatus.

(Fourth embodiment)

[0076] The fourth embodiment descries examples

ble to extract the characteristics of peak and valley of the spectral envelop of an input signal, and therefore to extract the static characteristics to detect a region with high possibility that the region is a speech region. Further, according to this constitution, it is possible to separate the speech region and stationary noise region with high accuracy.

[0093] First static characteristic extraction section 602 for quantized LSP parameter is comprised of components 614, 615 and 616 as described above.

[0094] In second static characteristic extraction section 603, reflective coefficient calculation section 617 converts the quantized LSP parameter into a reflective coefficient to output to voiced/unvoiced judgment section 620. Concurrently with the above processing, linear prediction residual power calculation section 618 calculates the linear prediction residual power from the quantized LSP parameter to output to voiced/unvoiced judgment section 620.

[0095] In addition, since linear prediction residual power calculation section 618 is the same as linear prediction residual power calculation section 614, it is possible to share one component as the sections 614 and 618.

[0096] Second static characteristic extraction section 603 for quantized LSP parameter is comprised of components 617 and 618 as described above.

[0097] Outputs from dynamic characteristic extraction section 601 and first static characteristic extraction section 602 are provided to speech region detection section 619. Speech region detection section 619 receives an evolution amount of the smoothed quantized LSP parameter input from square sum calculation section 607, a distance between the average quantized LSP parameter of the noise segment and the current quantized LSP parameter input from square sum calculation section 613, the quantized linear prediction residual power input from linear prediction residual power calculation section 614, and the variance information of the neighboring LSP region data input from variance calculation section 616. Then, using these information, speech region detection section 619 judges whether or not an input signal (or a decoded signal) at the current unit processing time is a speech region, and outputs the judged result to mode determination section 621. The more specific method for judging whether the input signal is a speech region is descried later using FIG.8.

[0098] On the other hand, an output from second characteristic extraction section 603 is provided to voiced/unvoiced judgment section 620. Voiced/unvoiced judgment section 620 receives the reflective coefficient input from reflective coefficient calculation section 617, and the quantized linear prediction residual power input from linear prediction residual power calculation section 618. Then, using these information, voiced/unvoiced judgment section 620 judges whether the input signal (decoded signal) at the current unit processing time is a voiced region or unvoiced region, and outputs the judged result to mode determination section 621. The more specific voiced/unvoiced judgment method is descried later using FIG.9.

[0099] Mode determination section 621 receives the judged result output from speech region detection section 619 and the judged result output from voiced/unvoiced judgment section 620, and using these information, determines a mode of the input signal (or decoded signal) at the current unit processing time to output. The more specific mode classifying method is described later using FIG.10.

[0100] In addition, although AR type sections are used as the smoothing section and average calculation section in this embodiment, it may be possible to perform the smoothing and average calculation by using other methods.

[0101] The detail of the speech region judgment method in the above-mentioned embodiment is next explained with reference to FIG.8.

[0102] First, in ST801, the first dynamic parameter (Paral) is calculated. The specific contents of the first dynamic parameter is an evolution amount of quantized LSP parameter for each unit processing time, and expressed with the following equation (3):

$$D(t)=\sum_{i=1}^{M}(LSi(t)-LSi(t-1))^2 \quad (3)$$

LSi(t): smoothed quantized LSP at time t

[0103] Next, in ST802, it is checked whether or not the first dynamic parameter is larger than a predetermined threshold Th1. When the parameter exceeds the threshold Th1, since the evolution amount of the quantized LSP parameter is large, it is judged that the input signal is a speech region. On the other hand, when the parameter is equal to or less than the threshold Th1, since the evolution amount of the quantized LSP parameter is small, the processing proceeds to ST803, and further proceeds to steps for judgment processing with other parameter.

[0104] In ST802, when the first dynamic parameter is equal to or less than the threshold Th1, the processing proceeds to ST803, where the number of a counter indicative of the number of times the stationary noise region is judged previously. The initial value of the counter is 0, and is incremented by 1 for each unit processing time judged as the stationary noise region with the mode determination method. In ST803, when the number of the counter equals to or less than a predetermined threshold ThC, the processing proceeds to ST804, where it is judged whether or not the input signal is a speech region using the static parameter. On the other hand, when the number of the counter exceeds the threshold ThC, the processing proceeds to ST806, where it is judged whether or not the input signal is a

coefficient exceeds the threshold Th2, the processing proceeds to ST905, and when the coefficient is equal to or less than the threshold Th2, the processing proceeds to ST904.

[0113] When the above-mentioned reflective coefficient is equal or less than the second threshold Th2 in ST903, in ST904, it is determined whether or not the above-mentioned reflective coefficient exceeds the third threshold Th3. When the coefficient exceeds the threshold Th3, the processing proceeds to ST907, and when the coefficient is equal to or less than the threshold Th3, the region is judged as the speech region, and the voiced/unvoiced judgment processing is finished.

[0114] When the above-mentioned reflective coefficient exceeds the second threshold Th2 in ST903, the linear prediction residual power is calculated in ST905. The linear prediction residual power is calculated after the quantized LSP is converted into the linear predictive coefficient.

[0115] In ST906, following ST905, it is determined whether or not the above-mentioned linear prediction residual power exceeds the threshold Th4. When the power exceeds the threshold Th4, it is judged that the region is the unvoiced region, and the voiced/unvoiced judgment processing is finished. When the power is equal to or less than the threshold Th4, it is judged that the region is the speech region, and the voiced/unvoiced judgment processing is finished.

[0116] When the above-mentioned reflective coefficient exceeds the third threshold Th3 in ST904, the linear prediction residual power is calculated in ST907.

[0117] In ST908, following ST907, it is determined whether or not the above-mentioned linear prediction residual power exceeds the threshold Th5. When the power exceeds the threshold Th5, it is judged that the region is the unvoiced region, and the voiced/unvoiced judgement processing is finished. When the power is equal to or less than the threshold Th5, it is judged that the region is the speech region, and the voiced/unvoiced judgment processing is finished.

[0118] The mode determination method used in mode determination section 621 is next explained with reference to FIG.10.

[0119] First, in ST1001, the speech region detection result is input. This step may be a block itself that performs the speech region detection processing.

[0120] Next, in ST1002, it is determined whether to determine that a mode is the stationary noise mode, based on the judgment result on whether or not the region is the speech region. When the region is the speech region, the processing proceeds to ST1003. When the region is not the speech region (stationary noise region), the mode determination result indicative of the stationary noise mode is output, and the mode determination processing is finished.

[0121] When it is determined that the region is not the stationary noise mode in ST1002, the voiced/unvoiced judgment result is input in ST1003. This step may be a block itself that performs the voiced/unvoiced determination processing.

[0122] Following ST1003, the mode determination is performed to determine whether the mode is the voiced region mode or the unvoiced region mode based on the voiced/unvoiced judgment result. When the judgment result is indicative of the voiced region, the mode determination result indicative of the voiced region mode is output, and the mode determination processing is finished. When the voiced/unvoiced judgment result is indicative of the unvoiced region, the mode determination result indicative of the unvoiced region mode is output, and the mode determination processing is finished. As described above, using the speech region detection result and voiced/unvoiced judgment, the modes of the input signals (or decoded signals) in a current unit processing block are classified into three modes.

(Fifth embodiment)

[0123] FIG.7 is a block diagram illustrating a configuration of a postprocessing section according to the fifth embodiment of the present invention. The postprocessing section is used in the speech signal decoding apparatus described in the second embodiment with the mode selector, described in the fourth embodiment, combined therewith. The postprocessing section illustrated in FIG.7 is provided with mode selection switches 705, 708, 707 and 711, spectral amplitude smoothing section 706, spectral phase randomizing sections 709 and 710, and threshold setting sections 703 and 716.

[0124] Weighted synthesis filter 701 receives decoded LPC output from LPC decoder 201 in the previously described speech decoding apparatus to construct the perceptual weighted synthesis filter, performs weighted filtering processing on the synthesized speech signal output from synthesis filter 209 or post filter 210 in the speech decoding apparatus to output to FFT processing section 702.

[0125] FFT processing section 702 performs FFT processing on the weighting-processed decoded signal output from weighted synthesis filter 701, and outputs a spectral amplitude WSAi to first threshold setting section 703, first spectral amplitude smoothing section 706 and first spectral phase randomizing section 709.

[0126] First threshold setting section 703 calculates the average of the spectral amplitude calculated in FFT processing section 702 using all frequency signal components, and using the calculated average as a reference, outputs the threshold Th1 to first spectral amplitude smoothing section 706 and first spectral phase randomizing section 709.

[0127] FFT processing section 704 performs FFT processing on the synthesized speech signal output from synthesis filter 209 and post filter 210 in the speech decoding apparatus, outputs the spectral amplitude to mode selection switches 705 and 712, adder 715, and second spectral phase randomizing section

processing section 720.

[0135] As mode selection switch 705, mode selection switch 712 receives the mode information (Mode) output from mode selector 202 in the speech decoding apparatus, and the difference information (Diff) output from adder 715, and judges whether the decoded signal in the current unit processing time is the speech region or the stationary noise region. When it is judged that the decoded signal is not the speech region (is the stationary noise region), mode selection switch 712 is connected to output the spectral amplitude SAi output from FFT processing section 704 to second spectral amplitude smoothing section 713. When it is determined that the decoded signal is the speech region, mode selection switch 712 is disconnected, and therefore the spectral amplitude SAi is not output to second spectral amplitude smoothing section 713.

[0136] Second spectral amplitude smoothing section 713 receives the spectral amplitude SAi output from FFT processing section 704 through mode selection switch 712, and performs the smoothing processing on signal components at all frequency bands. The average spectral amplitude in the stationary noise region can be obtained by this smoothing processing. The smoothing processing is the same as that in first spectral amplitude smoothing section 706. In addition, when mode selection switch 712 is disconnected, the section 713 does not perform the processing, and a smoothed spectral amplitude SSAi of the stationary noise region, which is last processed, is output. The smoothed spectral amplitude SSAi processed in second spectral amplitude smoothing processing section 713 is output to delay section 714, second threshold setting section 716, and mode selection switch 718.

[0137] Delay section 714 delays the input SSAi, output from second spectral amplitude smoothing section 713, by a unit processing time to output to adder 715.

[0138] Adder 715 calculates a difference between the smoothed spectral amplitude SSAi of the stationary noise region in the last unit processing time and the spectral amplitude SAi in the current unit processing time to output to mode switches 705, 707, 708, 711, 712, 718, and 719.

[0139] Second threshold setting section 716 sets the threshold Th2i using as a reference the smoothed spectral amplitude SSAi of the stationary noise region output from second spectral amplitude smoothing section 713 to output to second spectral phase randomizing section 710.

[0140] Random spectral phase generating section 717 outputs a randomly generated spectral phase to mode selection switch 719.

[0141] As mode selection switch 712, mode selection switch 718 receives the mode information (Mode) output from mode selector 202 in the speech decoding apparatus, and the difference information (Diff) output from adder 715, and judges whether the decoded signal in the current unit processing time is the speech region or the stationary noise region. When it is judged that the decoded signal is the speech region, mode selection switch 718 is connected to output an output from second spectral amplitude smoothing section 713 to IFFT processing section 720. When it is determined that the decoded signal is not the speech region (stationary noise region), mode selection switch 718 is disconnected, and therefore the output from second spectral amplitude smoothing section 713 is not output to IFFT processing section 720.

[0142] Mode selection switch 719 is switched synchronously with mode selection switch 718. As mode selection switch 718, mode selection switch 719 receives the mode information (Mode) output from mode selector 202 in the speech decoding apparatus, and the difference information (Diff) output from adder 715, and judges whether the decoded signal in the current unit processing time is the speech region or the stationary noise region. When it is judged that the decoded signal is the speech region, mode selection switch 719 is connected to output an output from random spectral phase generating section 717 to IIFFT processing section 720. When it is judged that the decoded signal is not the speech region (is stationary noise region), mode selection switch 719 is disconnected, and therefore the output from second random spectral phase generating section 717 is not output to IFFT processing section 720.

[0143] IFFT processing section 720 receives the spectral amplitude output from mode selection switch 707, the spectral phase output from mode selection switch 711, the spectral amplitude output from mode selection switch 718, and the spectral phase output from mode selection section 719 to perform IFFT processing, and outputs the processed signal. When mode selection switches 718 and 719 are disconnected, IFFT processing section 720 transforms the spectral amplitude input from mode selection 707 and the spectral phase input from mode selection switch 711 into a real part spectrum and imaginary part spectrum of FFT, then performs the IFFT processing, and outputs the real part of the resultant as a time signal. On the other hand, when mode selection switches 718 and 719 are connected, IFFT processing section 720 transforms the spectral amplitude input from mode selection 707 and the spectral phase input from mode selection switch 711 into a first real part spectrum and first imaginary part spectrum, and further transforms the spectral amplitude input from mode selection 718 and the spectral phase input from mode selection switch 719 into a second real part spectrum and second imaginary part spectrum to add, and then performs the IFFT processing. In other words, assuming that a third real part is obtained by adding the first real part spectrum to the second real part spectrum, and that a third imaginary part is obtained by adding the first imaginary part spectrum to the second imaginary part spectrum, the IFFT

ing is performed on the FFT spectral phase. The randomizing processing is performed on a signal component with a selected frequency in the same way as in the smoothing processing in ST1109. In other words, as in ST1109, the randomizing processing is performed on the signal component with the frequency i such that the perceptual weighted logarithmic spectral amplitude (WSAi) is equal to or less than the threshold Th1. At this point, it may be possible to set Th1 at the same value as in ST1109, and also possible to set Th1 at a different value adjusted to obtain higher subjective quality. In addition, random (i) in ST1110 is a numerical value ranging from -2π to $+2\pi$ generated randomly. To generate random (i), it may be possible to generate a random number newly every time. To save a computation amount, it may be also possible to hold pre-generated random numbers in a table to use while circulating the contents of the table for each unit processing time. When the table is used, two cases are considered that the contents of the table is used without modification, and that the contents of the table is added to the FFT spectral phase to use.

[0154] Next, in ST1111, a complex FFT spectrum is generated from the FFT logarithmic spectral amplitude and FFT spectral phase. The real part is obtained by returning the FFT logarithmic spectral amplitude SSA2i from the logarithmic region to the linear region, and then multiplying by a cosine of a spectral phase RSP2i. The imaginary part is obtained by returning the FFT logarithmic spectral amplitude SSA2i from the logarithmic region to the linear region, and then multiplying by a sine of the spectral phase RSP2i.

[0155] Next, in ST1112, the number of the counter indicative of the region judged as the stationary noise region is incremented by 1.

[0156] On the other hand, when it is judged that the decoded signal is the speech region (not the stationary noise region) in ST1106 or ST1107, next in ST1113, the FFT logarithmic spectral amplitude SAi is copied as the smoothed logarithmic spectrum SSA2i. In other words, the smoothing processing of the logarithmic spectral amplitude is not performed.

[0157] Next, in ST1114, the randomizing processing of the FFT spectral phase is performed. The randomizing processing is performed on a signal component with a selected frequency as in ST1110. However, the threshold for use in selecting the frequency is not Th1, but a value obtained by adding a constant k4 to SSAi previously obtained in ST1108. This threshold equals to the second threshold Th2i in FIG.6. In other words, the randomizing of the spectral phase is performed on a signal component with a frequency such that the spectral amplitude is smaller than the average spectral amplitude of the stationary noise region.

[0158] Next, in ST1115, a complex FFT spectrum is generated from the FFT logarithmic spectral amplitude and FFT spectral phase. The real part is obtained by adding the value obtained by returning the FFT logarithmic spectral amplitude SSA2i from the logarithmic region to the linear region, and then multiplying by the cosine of the spectral phase RSP2i, and a value obtained by multiplying a value obtained by returning the FFT logarithmic spectral amplitude SSAi from the logarithmic region to the linear region by a cosine of a spectral phase random2(i), and further multiplying the resultant by the constant k5. The imaginary part is obtained by adding the value obtained by returning the FFT logarithmic spectral amplitude SSA2i from the logarithmic region to the linear region, and then multiplying by the sine of the spectral phase RSP2i, and a value obtained by multiplying a value obtained by returning the FFT logarithmic spectral amplitude SSAi from the logarithmic region to the linear region by a sine of the spectral phase random2(i), and further multiplying the resultant by the constant k5. The constant k5 is in the range of 0.0 to 1.0, and specifically set at about 0.25. In addition, k5 may be an adaptively controlled variable. It is possible to improve the subjective qualities of the background stationary noise in the speech region by multiplexing the average stationary noise multiplied by k. The random2(i) is the same random number as random(i).

[0159] Next, in ST1116, IFFT is performed on complex FFT spectrum (Re(S2)i, Im(S2)i) generated in ST1111 or ST1115 to obtain a complex (Re(s2)i, Im(s2)i).

[0160] Finally, in ST1117, the real part Re(s2)i of the complex obtained by the IFFT is output.

[0161] According to the multimode speech coding apparatus of the present invention, since the coding mode of the second coding section is determined using the coded result in the first coding section, it is possible to provide the second coding section with the multimode without adding any new information indicative of a mode, and thereby to improve the coding performance.

[0162] In this constitution, the mode switching section switches the mode of the second coding section that encodes the excitation vector using the quantized parameter indicative of speech spectral characteristic, whereby in the speech coding apparatus that encodes parameters indicative of spectral characteristics and parameters indicative of the excitation vector independently of each other, it is possible to provide the coding of the excitation vector with the multimode without increasing new transmission information, and therefore to improve the coding performance.

[0163] In this case, since it is possible to detect the stationary noise segment using dynamic characteristics for the mode selection, the excitation vector coding provided with the multimode improves the coding performance for the stationary noise segment.

[0164] Further, in this case, the mode switching section switches the mode of the processing section that encodes the excitation vector using quantized LSP parameters, and therefore it is possible to apply the present invention simply to a CELP system that uses

ing:

first decoding means for decoding at least one type of parameter indicative of vocal tract information contained in a speech signal;
second decoding means for being capable of decoding said at least one type of parameter indicative of vocal tract information with a plurality of decoding modes;
mode switching means for switching a decoding mode of said second decoding means based on a dynamic characteristic of a specific parameter decoded in said first decoding means; and
synthesis means for decoding the speech signal using a plurality of types of parameter information decoded in said first decoding means and said second decoding means.

8. The multimode speech decoding apparatus according to claim 7, wherein said second decoding means comprises decoding means for being capable of decoding an excitation vector with a plurality of decoding modes, and said mode switching means switches the decoding mode of said second decoding means using a quantized parameter indicative of a spectral characteristic of a speech.

9. The multimode speech decoding apparatus according to claim 8, wherein said mode switching means switches the decoding mode of said second decoding means using a static characteristic and a dynamic characteristic of the quantized parameter indicative of the spectral characteristic of the speech.

10. The multimode speech decoding apparatus according to claim 8, wherein said mode switching means switches the decoding mode of said second decoding means using a quantized LSP parameter.

11. The multimode speech decoding apparatus according to claim 10, wherein said mode switching means switches the decoding mode of said second decoding means using a static characteristic and a dynamic characteristic of the quantized LSP parameter.

12. The multimode speech decoding apparatus according to claim 10, wherein said mode switching means comprises means for judging stationarity of the quantized LSP parameter using a previous quantized LSP parameter and a current quantized LSP parameter, and means for judging a voiced characteristic using the current quantized LSP parameter, and based on judged results, switches the decoding mode of said second decoding means.

13. The multimode speech decoding apparatus according to claim 7, wherein said apparatus switches postprocessing for a decoded signal based on judged results.

14. A quantized-LSP-parameter dynamic characteristic extractor comprising:

means for calculating an evolution of a quantized LSP parameter between frames;
means for calculating an average quantized LSP parameter in a frame in which the quantized LSP parameter is stationary; and
means for calculating an evolution between said average quantized LSP parameter and a current quantized LSP parameter.

15. A quantized-LSP-parameter static characteristic extractor comprising:

means for calculating linear prediction residual power using a quantized LSP parameter; and
means for calculating a region between neighboring orders of the quantized LSP parameter.

16. A multimode postprocessing apparatus comprising:

judgment means for judging whether or not a region is a speech region using a decoded LSP parameter;
FFT processing means for performing fast Fourier transform processing on a signal;
spectral phase randomizing means for randomizing a spectral phase obtained by said fast Fourier transform processing corresponding to a result judged by said judgment means;
spectral amplitude smoothing means for performing smoothing on a spectral amplitude obtained by said fast Fourier transform processing corresponding to said result; and
IFFT processing means for performing inverse fast Fourier transform on the spectral phase randomized by said spectral phase randomizing means and the spectral amplitude smoothed by said spectral amplitude smoothing means.

17. The multimode postprocessing apparatus according to claim 16, wherein said device determines a frequency of the spectral phase to be randomized using an average spectral amplitude of a previous non-speech region in a speech region, and determines a frequency of the spectral phase to be randomized and the spectral amplitude to be smoothed using an average spectral amplitude with all frequencies in a perceptual weighted domain in a non-speech region.

extracting method comprising the steps of:

calculating an evolution of a quantized LSP parameter between frames;
calculating an average quantized LSP parameter in a frame in which the quantized LSP parameter is stationary; and
calculating an evolution between said average quantized LSP parameter and a current quantized LSP parameters.

27. A quantized-LSP-parameter static characteristic extracting method comprising the steps:

calculating linear prediction residual power using a quantized LSP parameter; and
calculating a region between neighboring orders of the quantized LSP parameter.

28. A multimode postprocessing method comprising:

the judgment step of judging whether or not a region is a speech region using a decoded LSP parameter;
the FFT processing step of performing fast Fourier transform processing on a signal;
the spectral phase randomizing step of randomizing a spectral phase obtained by said fast Fourier transform processing corresponding to a result determined by said judgment step;
the spectral amplitude smoothing step of performing smoothing on a spectral amplitude obtained by said fast Fourier transform processing corresponding to said result; and
the IFFT processing step of performing inverse fast Fourier transform on the spectral phase randomized by said spectral phase randomizing step and the spectral amplitude smoothed by said spectral amplitude smoothing step.

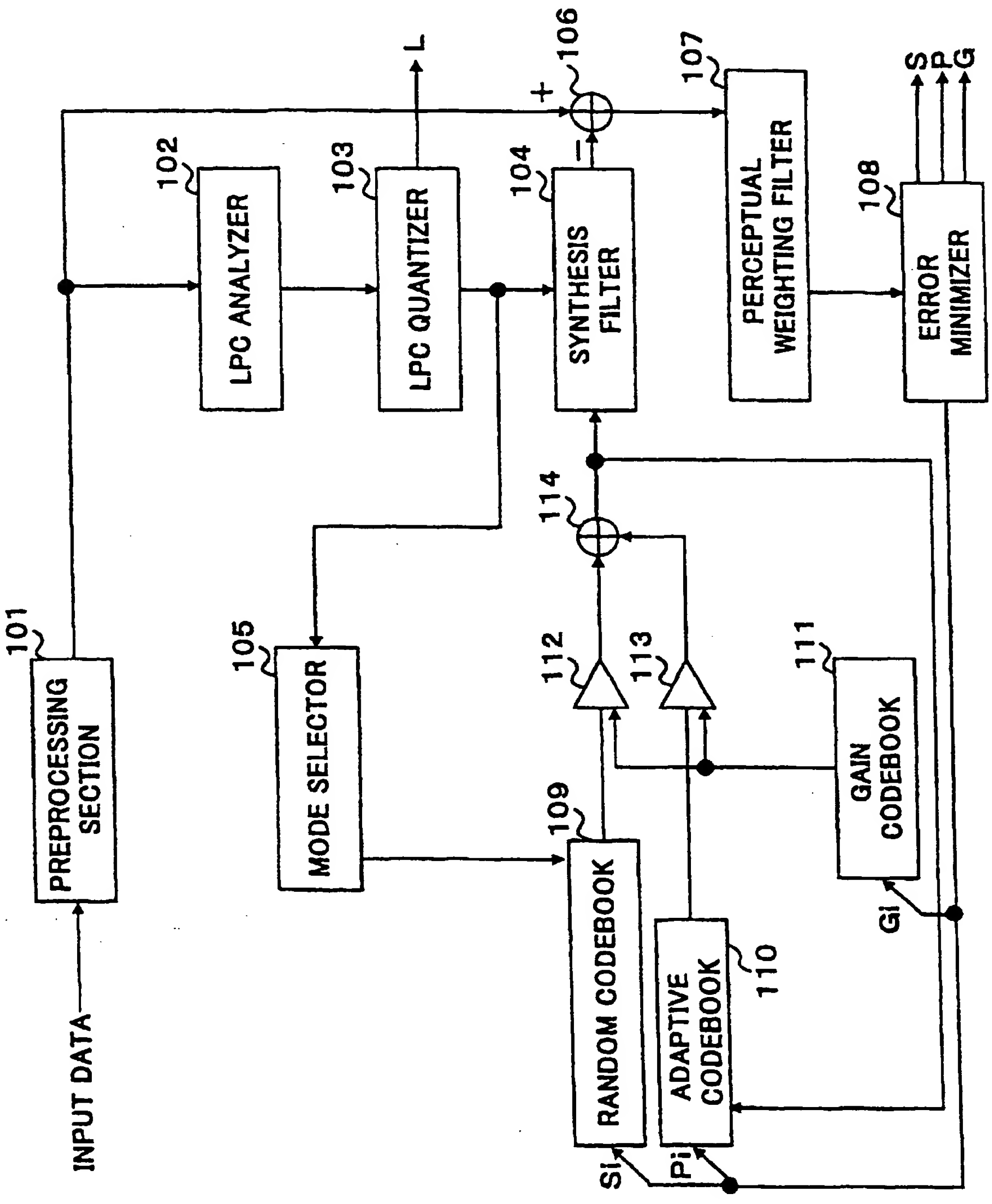
INPUT DATA
101
PREPROCESSING SECTION
102
LPC ANALYZER
103
LPC QUANTIZER
L
104
SYNTHESIS FILTER
+
−
106
107
PERCEPTUAL WEIGHTING FILTER
108
ERROR MINIMIZER
S
P
G
105
MODE SELECTOR
114
112
113
109
RANDOM CODEBOOK
110
ADAPTIVE CODEBOOK
111
GAIN CODEBOOK
Si
Pi
Gi

FIG.1

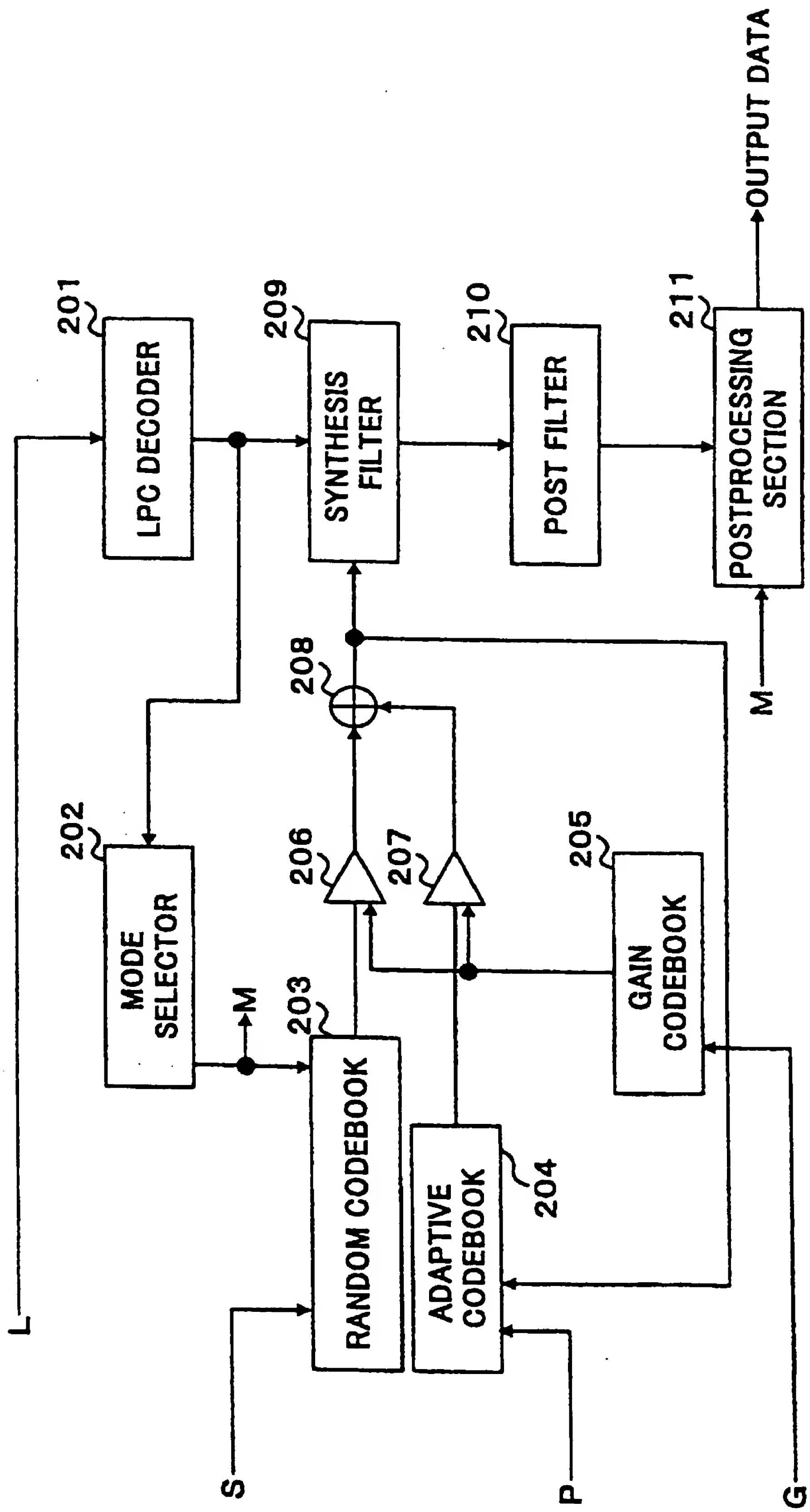
L
S
P
G
M
201
LPC DECODER
202
MODE SELECTOR
203
RANDOM CODEBOOK
204
ADAPTIVE CODEBOOK
205
GAIN CODEBOOK
206
207
208
209
SYNTHESIS FILTER
210
POST FILTER
211
POSTPROCESSING SECTION
OUTPUT DATA

FIG.2

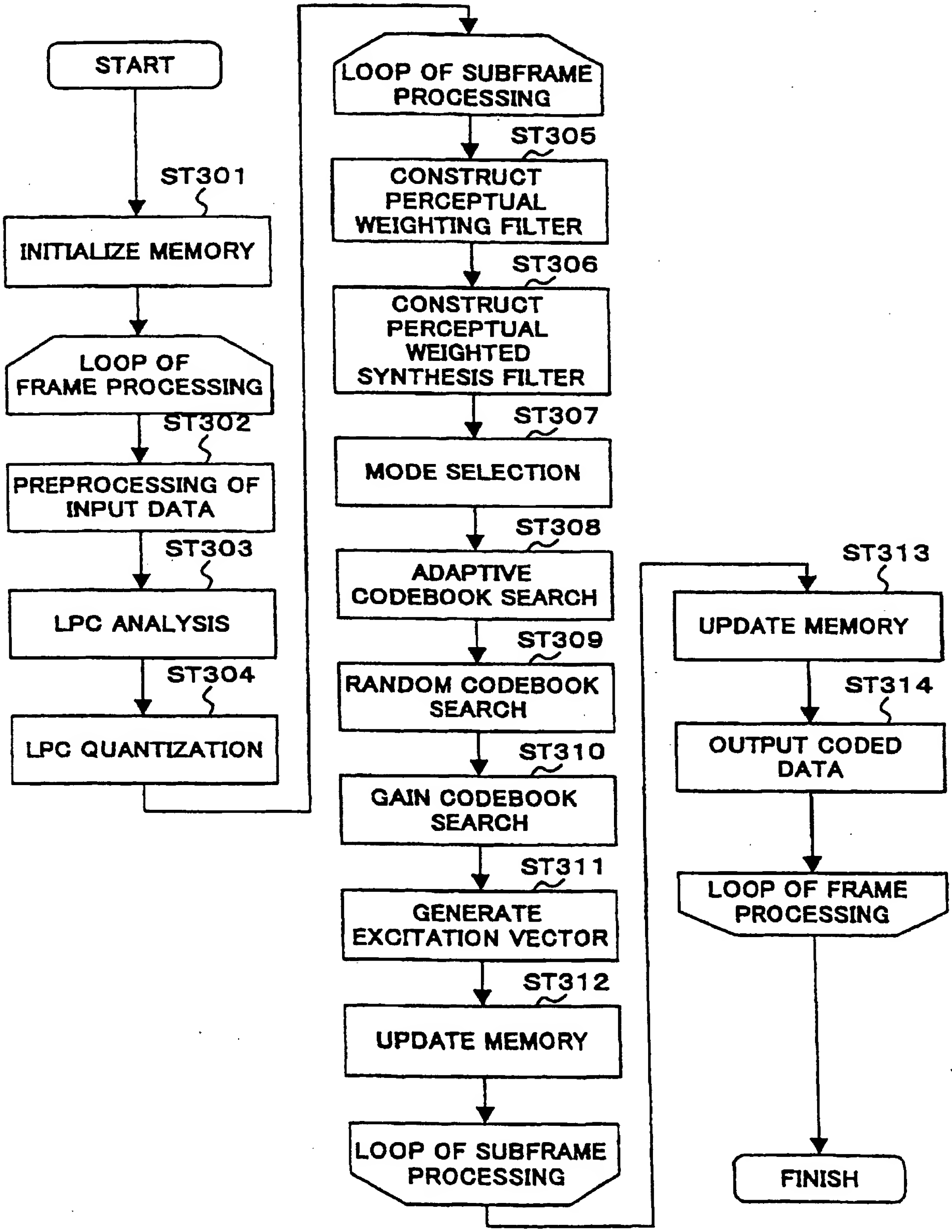
START
ST301
INITIALIZE MEMORY
LOOP OF FRAME PROCESSING
ST302
PREPROCESSING OF INPUT DATA
ST303
LPC ANALYSIS
ST304
LPC QUANTIZATION
LOOP OF SUBFRAME PROCESSING
ST305
CONSTRUCT PERCEPTUAL WEIGHTING FILTER
ST306
CONSTRUCT PERCEPTUAL WEIGHTED SYNTHESIS FILTER
ST307
MODE SELECTION
ST308
ADAPTIVE CODEBOOK SEARCH
ST309
RANDOM CODEBOOK SEARCH
ST310
GAIN CODEBOOK SEARCH
ST311
GENERATE EXCITATION VECTOR
ST312
UPDATE MEMORY
LOOP OF SUBFRAME PROCESSING
ST313
UPDATE MEMORY
ST314
OUTPUT CODED DATA
LOOP OF FRAME PROCESSING
FINISH

FIG.3

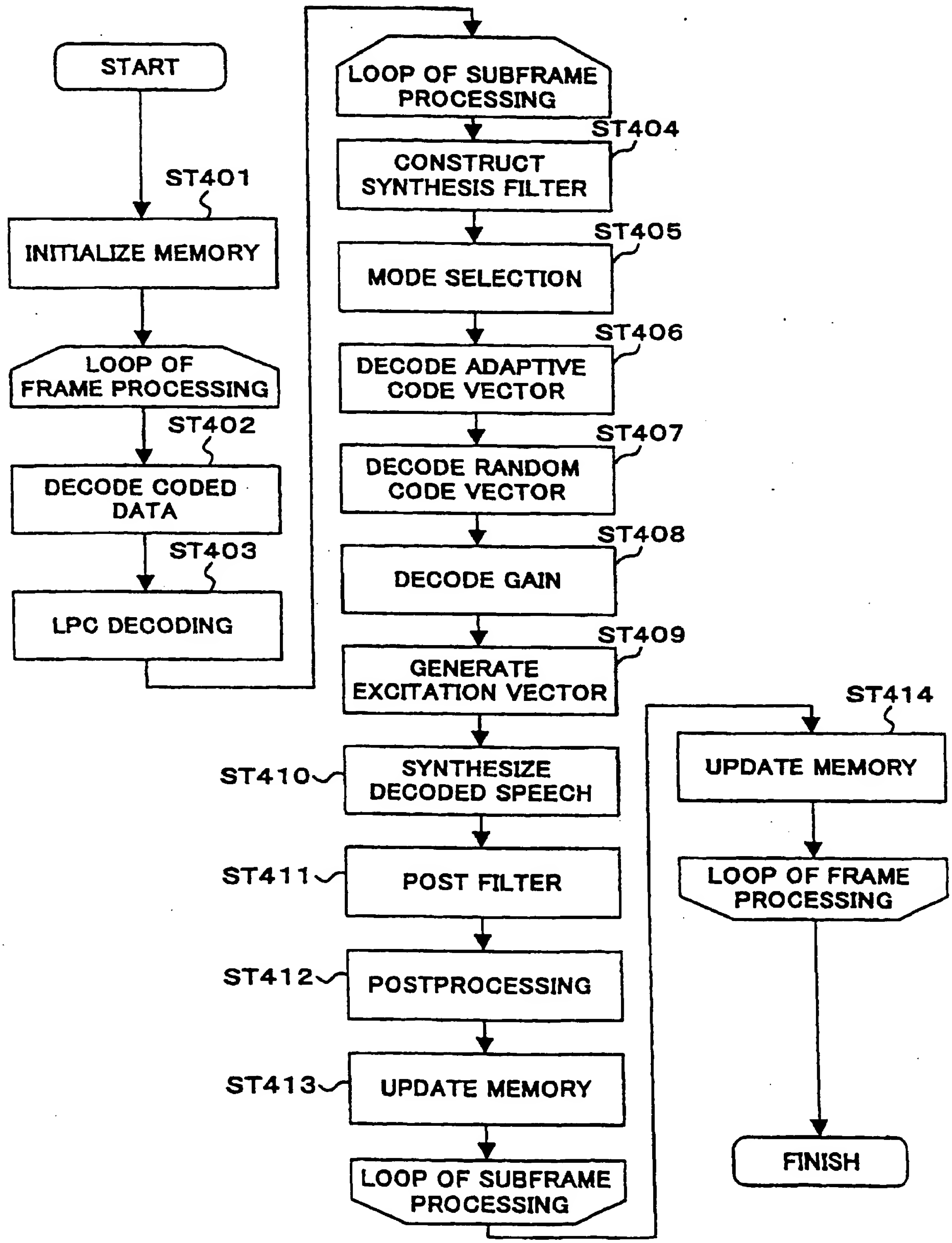
START
ST401
INITIALIZE MEMORY
LOOP OF FRAME PROCESSING
ST402
DECODE CODED DATA
ST403
LPC DECODING
LOOP OF SUBFRAME PROCESSING
ST404
CONSTRUCT SYNTHESIS FILTER
ST405
MODE SELECTION
ST406
DECODE ADAPTIVE CODE VECTOR
ST407
DECODE RANDOM CODE VECTOR
ST408
DECODE GAIN
ST409
GENERATE EXCITATION VECTOR
ST410
SYNTHESIZE DECODED SPEECH
ST411
POST FILTER
ST412
POSTPROCESSING
ST413
UPDATE MEMORY
LOOP OF SUBFRAME PROCESSING
ST414
UPDATE MEMORY
LOOP OF FRAME PROCESSING
FINISH

FIG.4

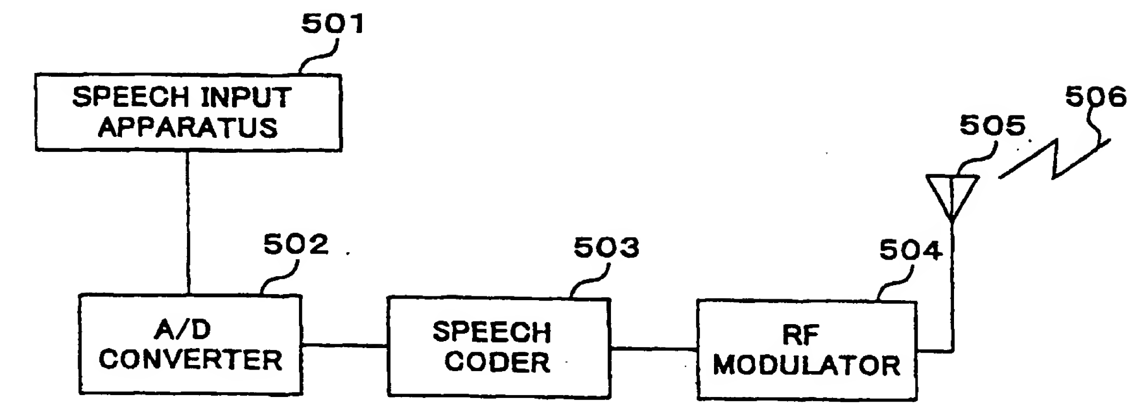
501
SPEECH INPUT APPARATUS
502
A/D CONVERTER
503
SPEECH CODER
504
RF MODULATOR
505
506

FIG.5A

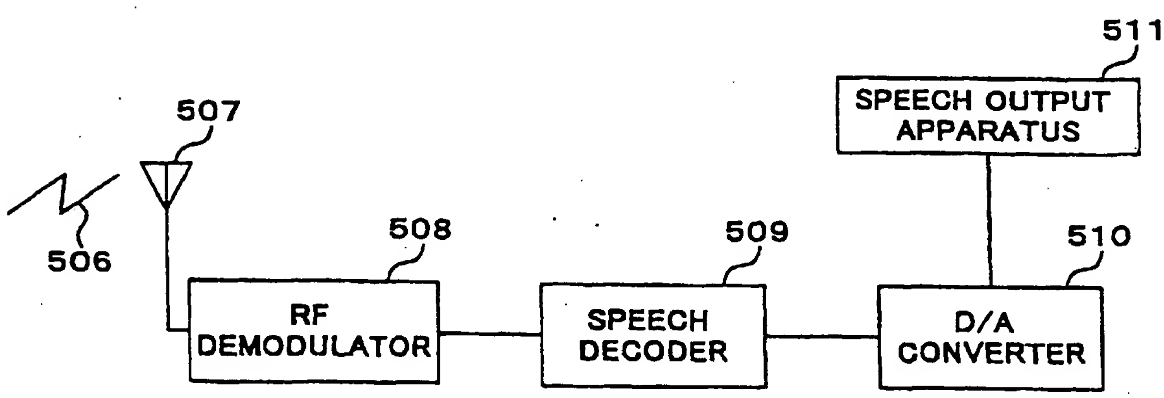
511
SPEECH OUTPUT APPARATUS
507
506
508
RF DEMODULATOR
509
SPEECH DECODER
510
D/A CONVERTER

FIG.5B

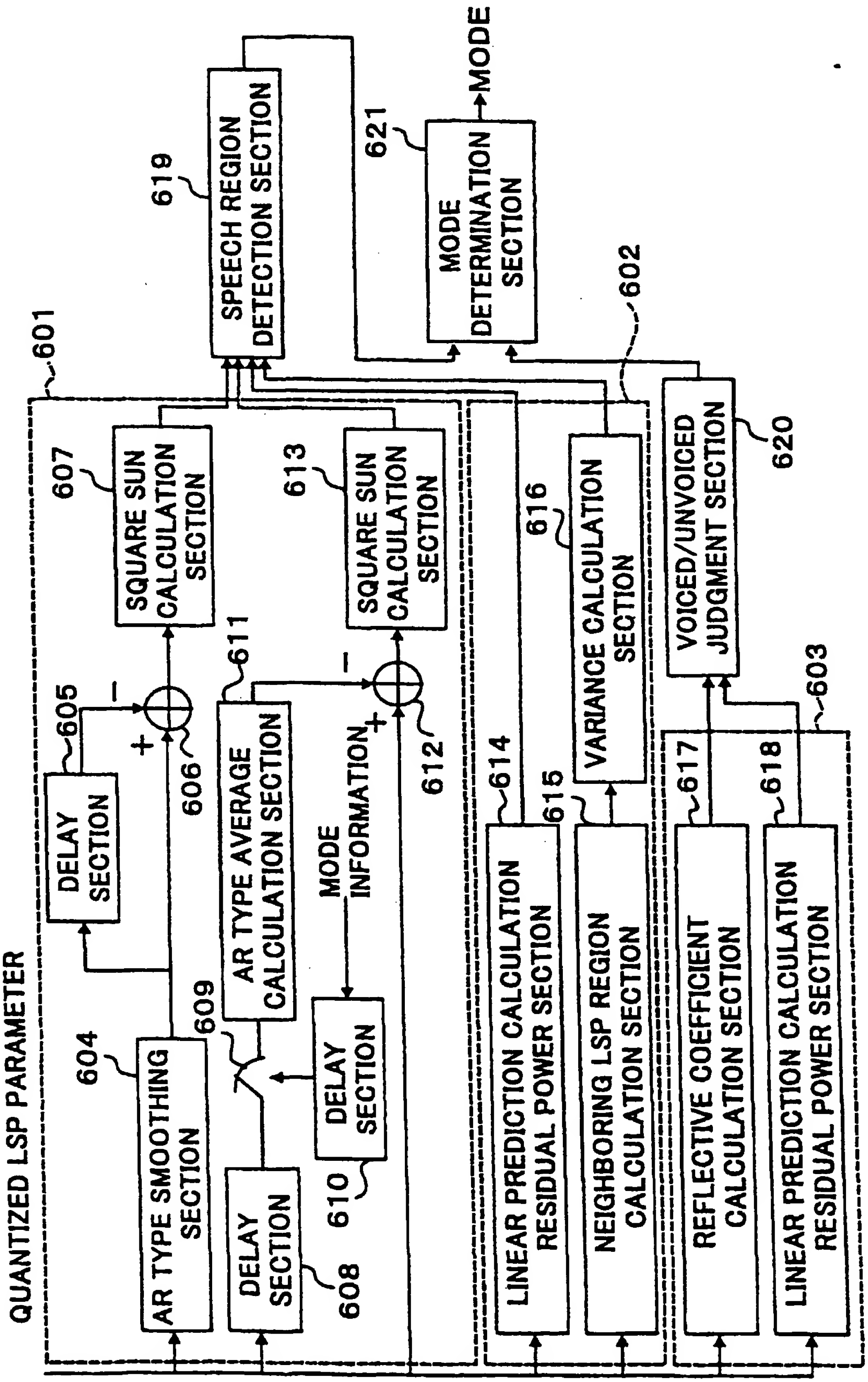
QUANTIZED LSP PARAMETER
601
AR TYPE SMOOTHING SECTION
604
DELAY SECTION
605
606
SQUARE SUN CALCULATION SECTION
607
DELAY SECTION
608
609
DELAY SECTION
610
AR TYPE AVERAGE CALCULATION SECTION
611
MODE INFORMATION
612
SQUARE SUN CALCULATION SECTION
613
LINEAR PREDICTION CALCULATION RESIDUAL POWER SECTION
614
NEIGHBORING LSP REGION CALCULATION SECTION
615
VARIANCE CALCULATION SECTION
616
602
REFLECTIVE COEFFICIENT CALCULATION SECTION
617
LINEAR PREDICTION CALCULATION RESIDUAL POWER SECTION
618
603
VOICED/UNVOICED JUDGMENT SECTION
620
SPEECH REGION DETECTION SECTION
619
MODE DETERMINATION SECTION
621
MODE

FIG.6

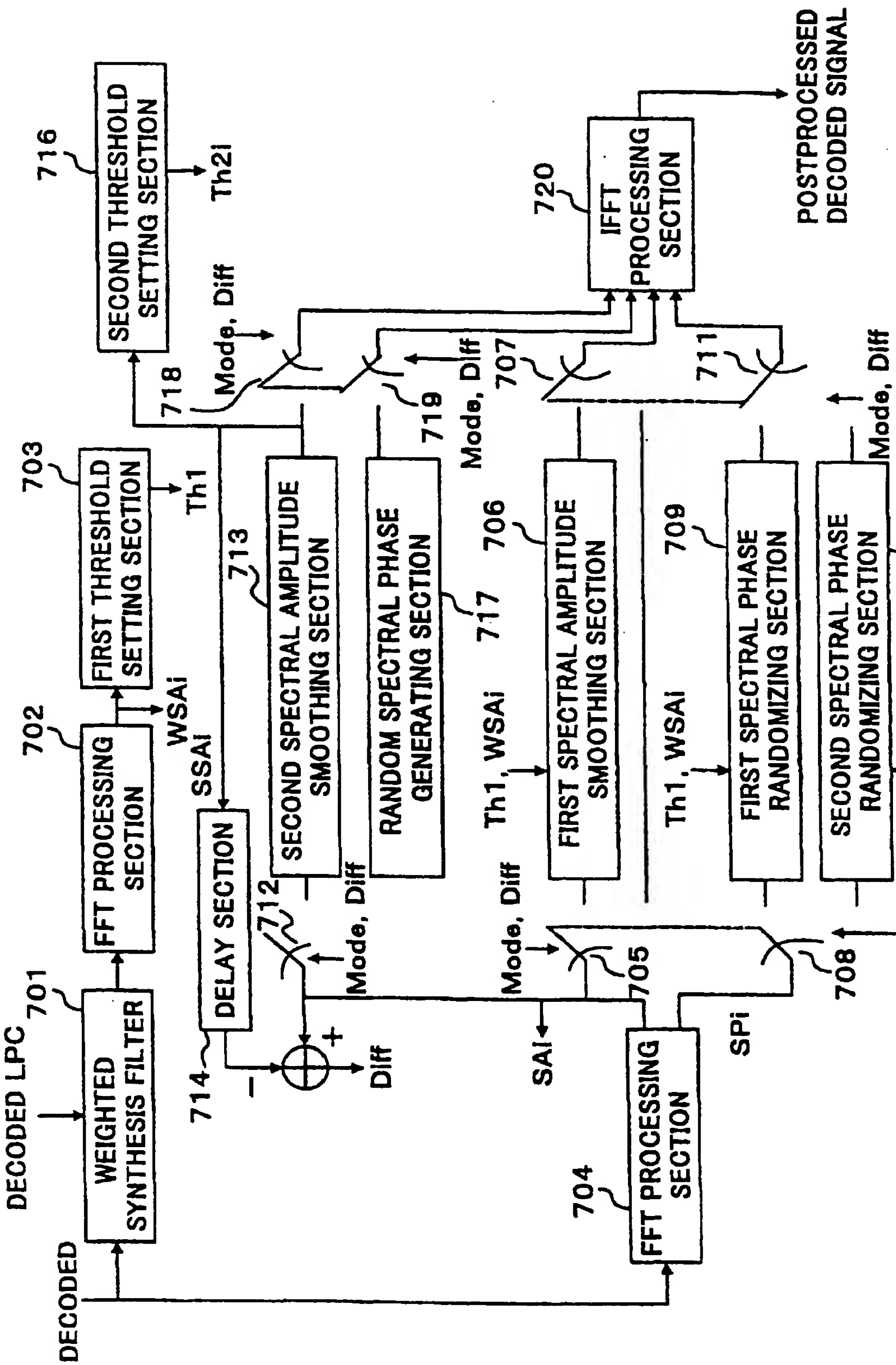
DECODED LPC
DECODED
701
WEIGHTED SYNTHESIS FILTER
702
FFT PROCESSING SECTION
WSAi
703
FIRST THRESHOLD SETTING SECTION
Th1
716
SECOND THRESHOLD SETTING SECTION
Th2i
SSAi
714
DELAY SECTION
Diff
712
Mode, Diff
713
SECOND SPECTRAL AMPLITUDE SMOOTHING SECTION
717
RANDOM SPECTRAL PHASE GENERATING SECTION
718
719
SAi
SPi
704
FFT PROCESSING SECTION
705
706
Th1, WSAi
FIRST SPECTRAL AMPLITUDE SMOOTHING SECTION
707
708
709
FIRST SPECTRAL PHASE RANDOMIZING SECTION
710
Th2i, SAi
SECOND SPECTRAL PHASE RANDOMIZING SECTION
711
720
IFFT PROCESSING SECTION
POSTPROCESSED DECODED SIGNAL

FIG.7

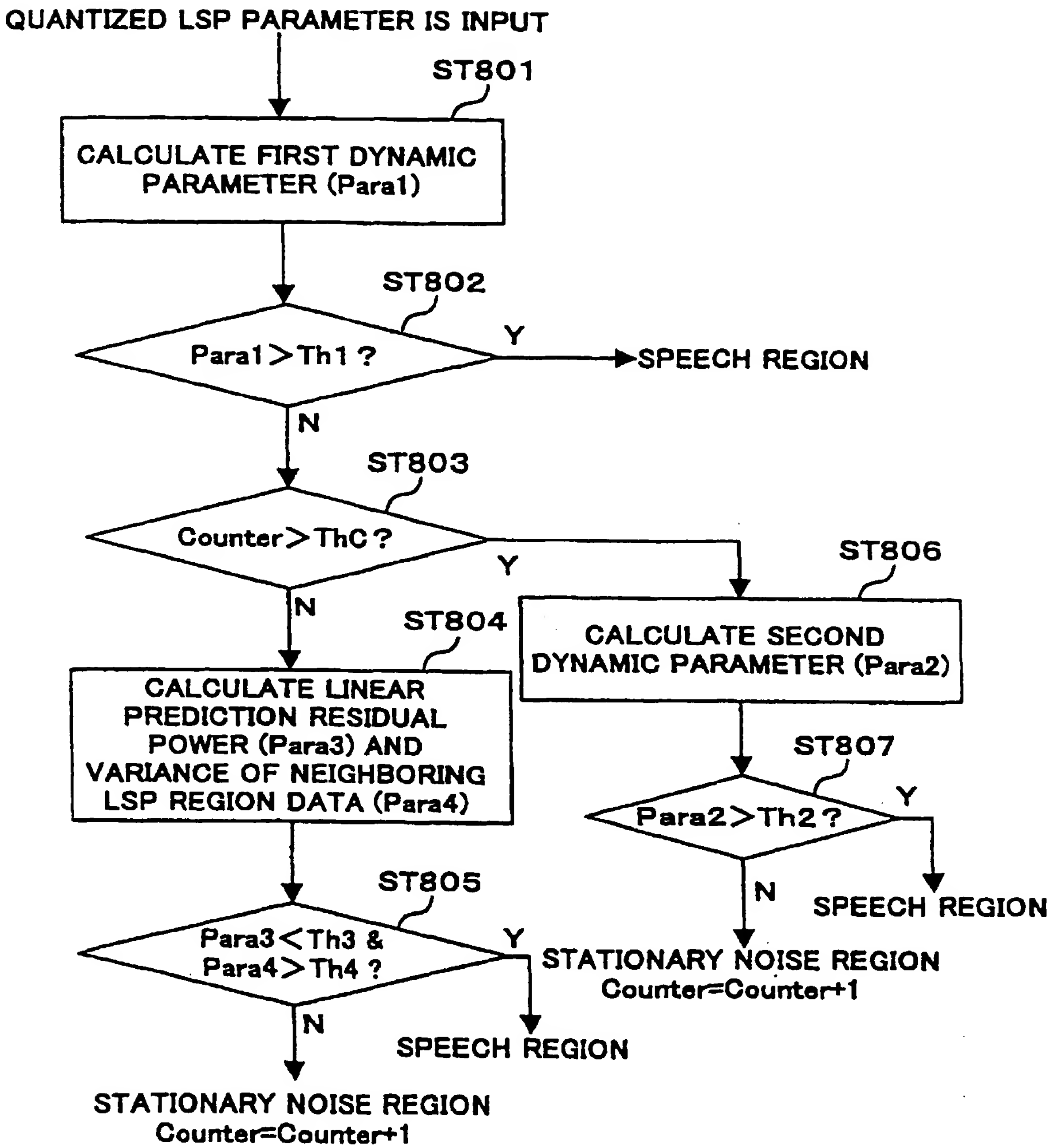
QUANTIZED LSP PARAMETER IS INPUT
ST801
CALCULATE FIRST DYNAMIC PARAMETER (Para1)
ST802
Para1>Th1 ?
Y
SPEECH REGION
N
ST803
Counter>ThC ?
Y
ST806
CALCULATE SECOND DYNAMIC PARAMETER (Para2)
N
ST804
CALCULATE LINEAR PREDICTION RESIDUAL POWER (Para3) AND VARIANCE OF NEIGHBORING LSP REGION DATA (Para4)
ST807
Para2>Th2 ?
Y
N
SPEECH REGION
STATIONARY NOISE REGION
Counter=Counter+1
ST805
Para3<Th3 &
Para4>Th4 ?
Y
N
SPEECH REGION
STATIONARY NOISE REGION
Counter=Counter+1

FIG.8

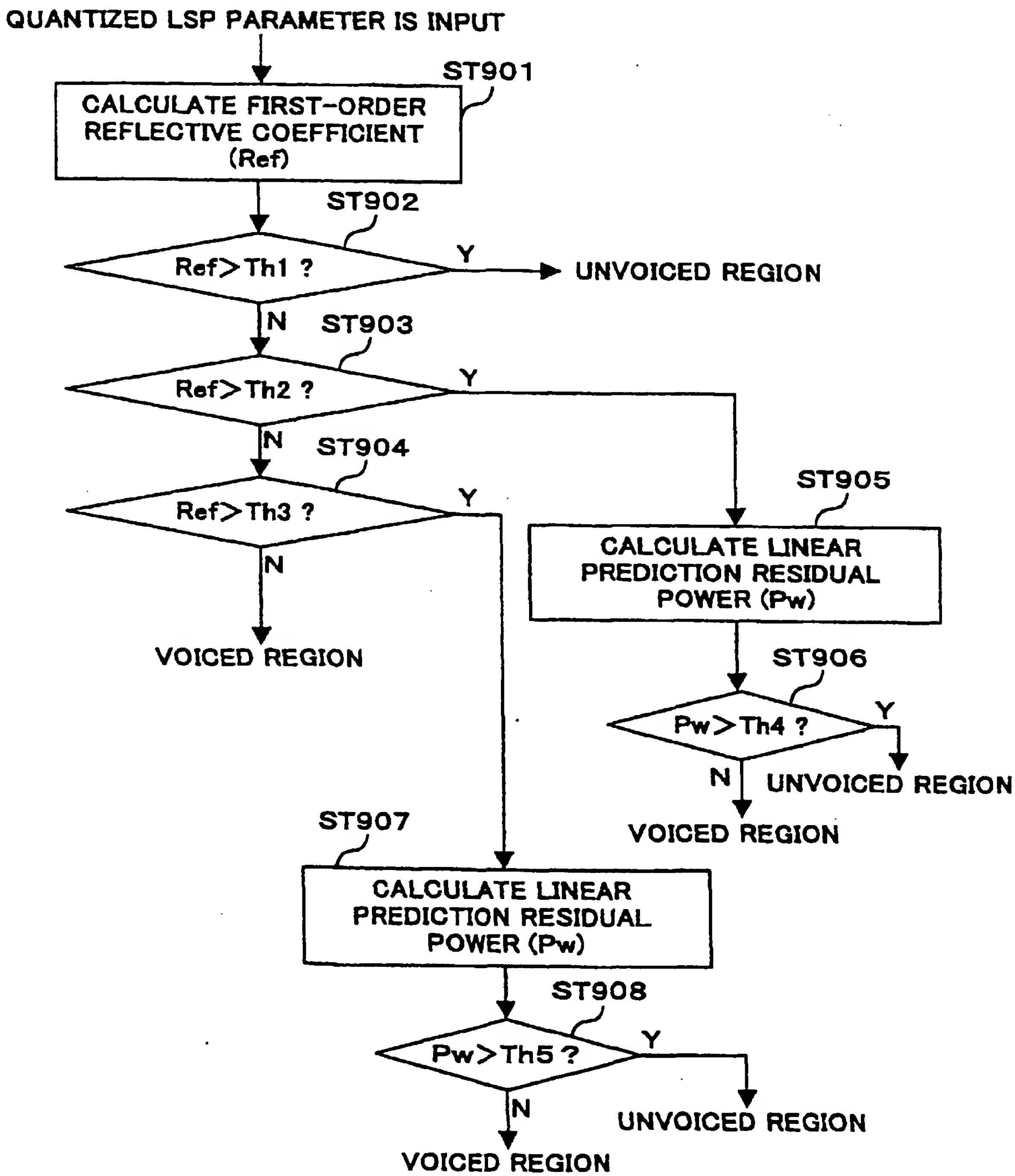
QUANTIZED LSP PARAMETER IS INPUT
ST901
CALCULATE FIRST-ORDER REFLECTIVE COEFFICIENT (Ref)
ST902
Ref>Th1 ?
Y
UNVOICED REGION
N
ST903
Ref>Th2 ?
Y
N
ST904
Ref>Th3 ?
Y
N
VOICED REGION
ST905
CALCULATE LINEAR PREDICTION RESIDUAL POWER (Pw)
ST906
Pw>Th4 ?
Y
N
UNVOICED REGION
VOICED REGION
ST907
CALCULATE LINEAR PREDICTION RESIDUAL POWER (Pw)
ST908
Pw>Th5 ?
Y
N
UNVOICED REGION
VOICED REGION

FIG.9

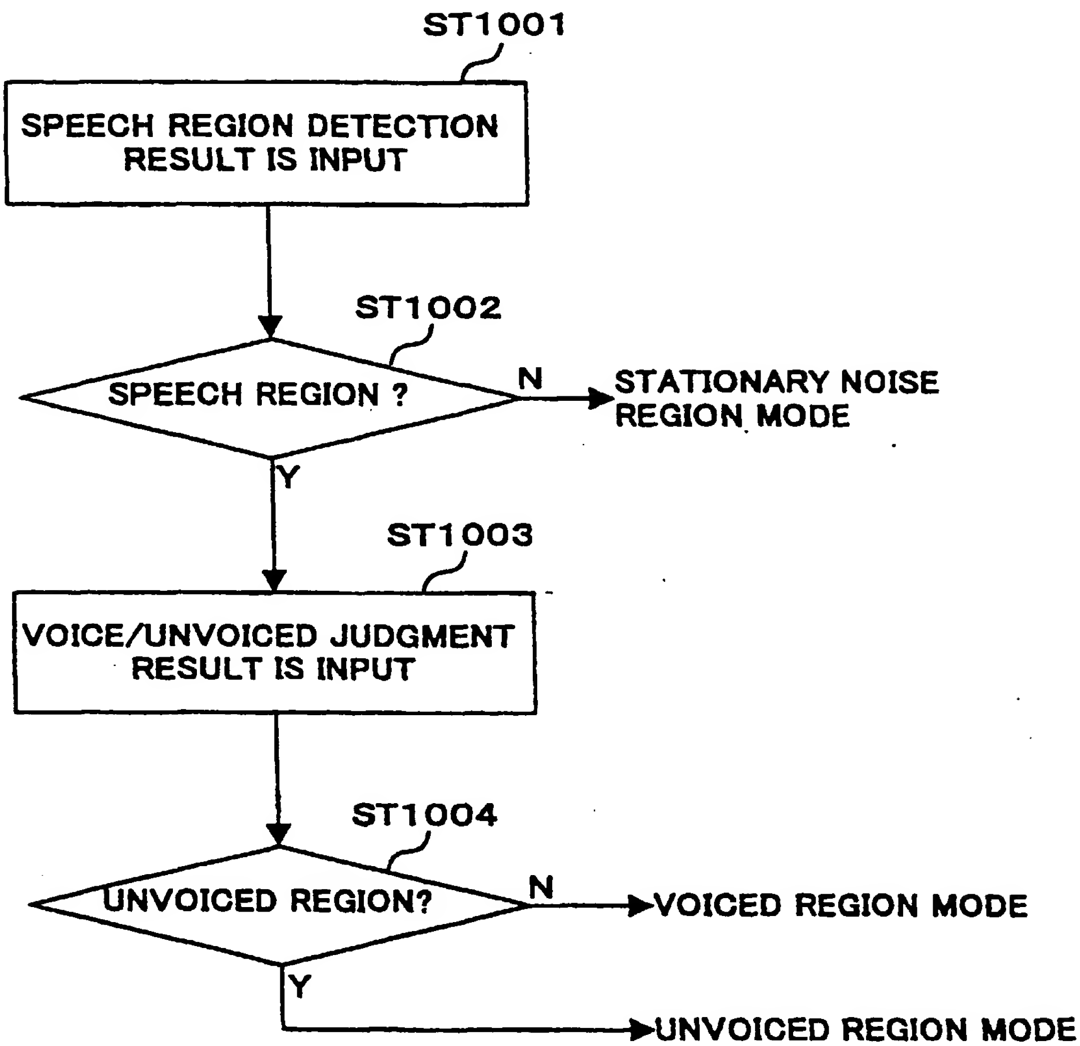
ST1001
SPEECH REGION DETECTION RESULT IS INPUT
ST1002
SPEECH REGION ?
N
STATIONARY NOISE REGION MODE
Y
ST1003
VOICE/UNVOICED JUDGMENT RESULT IS INPUT
ST1004
UNVOICED REGION?
N
VOICED REGION MODE
Y
UNVOICED REGION MODE

FIG.10

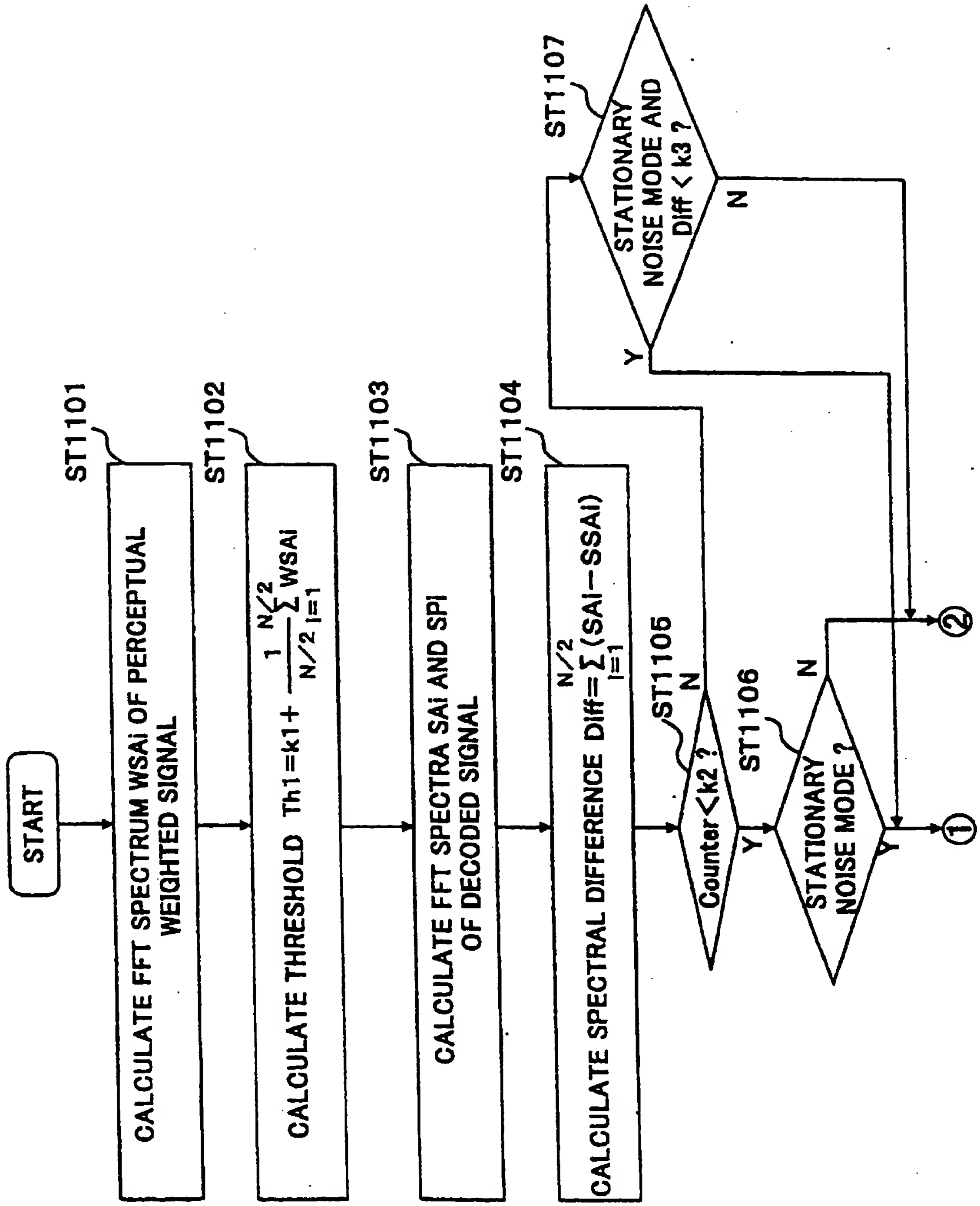
START
ST1101
CALCULATE FFT SPECTRUM WSAi OF PERCEPTUAL WEIGHTED SIGNAL
ST1102
CALCULATE THRESHOLD $Th1=k1+\frac{1}{N/2}\sum_{i=1}^{N/2} WSAi$
ST1103
CALCULATE FFT SPECTRA SAi AND SPi OF DECODED SIGNAL
ST1104
CALCULATE SPECTRAL DIFFERENCE $Diff=\sum_{i=1}^{N/2}(SAi-SSAi)$
ST1105
Counter<k2 ?
Y
N
ST1106
STATIONARY NOISE MODE ?
Y
N
ST1107
STATIONARY NOISE MODE AND Diff < k3 ?
Y
N
1
2

FIG.11

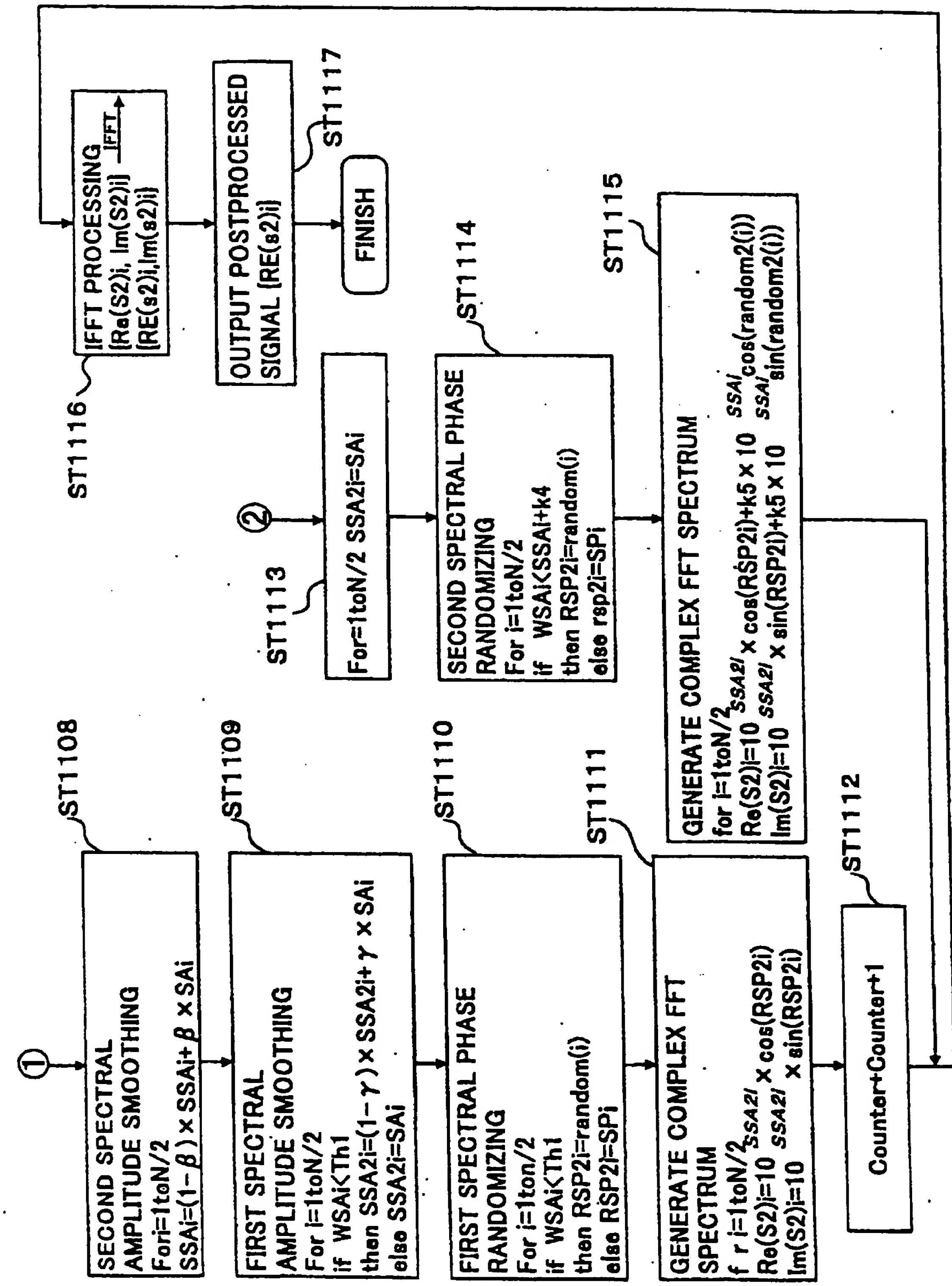
①
ST1108
SECOND SPECTRAL AMPLITUDE SMOOTHING
Fori=1toN/2
SSAi=(1-β)×SSAi+β×SAi
ST1109
FIRST SPECTRAL AMPLITUDE SMOOTHING
For i=1toN/2
if WSAi<Th1
then SSA2i=(1-γ)×SSA2i+γ×SAi
else SSA2i=SAi
ST1110
FIRST SPECTRAL PHASE RANDOMIZING
For i=1ton/2
if WSAi<Th1
then RSP2i=random(i)
else RSP2i=SPi
ST1111
GENERATE COMPLEX FFT SPECTRUM
f r i=1toN/2
Re(S2)i=10 SSA2i × cos(RSP2i)
Im(S2)i=10 SSA2i × sin(RSP2i)
ST1112
Counter+Counter+1
②
ST1113
For=1toN/2 SSA2i=SAi
ST1114
SECOND SPECTRAL PHASE RANDOMIZING
For i=1toN/2
if WSAi<SSAi+k4
then RSP2i=random(i)
else rsp2i=SPi
ST1115
GENERATE COMPLEX FFT SPECTRUM
for i=1toN/2
Re(S2)i=10 SSA2i × cos(RSP2i)+k5×10 SSAi cos(random2(i))
Im(S2)i=10 SSA2i × sin(RSP2i)+k5×10 SSAi sin(random2(i))
ST1116
IFFT PROCESSING
[Re(S2)i, Im(S2)i] IFFT
[RE(s2)i,Im(s2)i]
OUTPUT POSTPROCESSED SIGNAL [RE(s2)i]
ST1117
FINISH

FIG.12